ISSN print: 1224-5984
ISSN online: 2248-3942

Rom Biotechnol Lett. 2021; 26(4): 2737-2750
doi: 10.25083/rbl/26.4/2737.2750

Received for publication, March, 8, 2021
Accepted, April, 5, 2021

Original paper

Comparative Computational Analysis of SARS-CoV-2 Nucleocapsid Epitope with Taxonomically Related Corona Viruses

SHEREEN F. ELKHOLY[1,2]

[1]Biology Department, College of Science and Humanities, Prince Sattam bin Abdulaziz University, 11942 Alkharj, Kingdom of Saudi Arabia
[2]Department of Plant Molecular Biology, Plant Transformation and Biopharmaceuticals Lab, Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Centre (ARC), Giza, Egypt

Abstract

Several research lines are currently ongoing to address the multitude of facets of the pandemic COVID-19. In line with the One-Health concept, extending the target of the studies to the animals which humans are continuously interacting with may favor a better understanding of the SARS-CoV-2 Biology and pathogenetic mechanisms; thus, helping to adopt the most suitable containment measures. The last two decades have already faced severe manifestations of the coronavirus infection in both humans and animals, thus, circulating epitopes from previous outbreaks might confer partial protection from SARS-CoV-2 infections. In the present study, we provide computational analysis of the major nucleocapsid protein epitopes and compare them with the homologues of taxonomically-related coronaviruses with tropism for animal species that are closely inter-related with the human being population all over the world. Protein sequence alignment provides evidence of high sequence homology for some of the investigated proteins. Moreover, the way the receptor binding domains of the nucleocapsid epitopes interact with their specific proteins is different from the closely related viruses. These evidences provide a molecular structural rationale for a potential role in conferring protection from SARS-CoV-2 infection and identifying potential candidates for the development of diagnostic tools and prophylactic- oriented strategies.

Keywords Bioinformatics, COVID-19, SARS-COV-2, nucleocapsid (NC), epitopes.

To cite this article: ELKHOLY SF. Comparative Computational Analysis of SARS-CoV-2 Nucleocapsid Epitope with Taxonomically Related Corona Viruses. *Rom Biotechnol Lett*. 2021; 26(4): 2737-2750. DOI: 10.25083/rbl/26.4/2737-2750

*Corresponding author: SHEREEN F. ELKHOLY, Biology Department, College of Science and Humanities, Prince Sattam bin Abdulaziz University, 11942 Alkharj, Kingdom of Saudi Arabia
E-mail: s.elkholy@psau.edu.sa

Introduction

The novel coronavirus, known as SARS-CoV-2, was identified as the cause of an outbreak of pneumonia in Wuhan, China, in December 2019 (XIE, M & CHEN 2020). Cases of coronavirus disease 2019 (COVID-19) associated with travel outside of China were reported as early as January 13, 2020, and the virus has subseq-uently spread to nearly all countries (World Health Organization, 2020).

Coronavirus is a large group of viruses that can infect animals and humans alike, causing respiratory diseases, whether they are mild like the common cold or severe such as pneumonia. Animal coronaviruses rarely infect humans and then spread among them. And we might remember SARS (severe acute respiratory syndrome), which spread between 2002-2003, and it is an example of the Corona virus that was transmitted from animals to humans. Another prominent modern Corona virus called MERS (Middle East Respiratory Syndrome) (CHINSEMBU et al, 2020) appeared in the Middle East in 2012, and scientists say it initially passed from a camel to a human. Coronavirus (SARS-CoV) is a type of Corona viruses that infect humans, bats (CHINSEMBU et al, 2020).

The WHO confirmed that the outbreak of the coronavirus epidemic was associated with the Huanan South China Seafood Marketplace, but no specific animal association was identified. Scientists immediately started to research the source of the new coronavirus, and the first genome of COVID-19 was published by the research team led by Prof. Yong-Zhen Zhang, on 10 January 2020 (ADHIKARI et al, 2020). Additionally, World Health Organization (WHO) has classified COVID-19 as a β CoV of cluster 2B.

Ten order sequences of COVID-19 obtained from a complete of 9 patients exhibited 99.98% sequence identity. Another study showed there was 99.8-99.9% nucleotide identity in isolates from 5 patients and also the sequence results. Disclosed the presence of a replacement beta-CoV strain The genetic sequence of the COVID-19 showed a lot of than 80% identity to SARS-CoV and 50% to the MERS-CoV, and each SARS-CoV and MERS-CoV originate in bats (HUSSIN et al 2020).

Coronaviruses are single-stranded, positive-sense, RNA viruses belonging to the Nidovirales order and were divided into 4 vast groups (i.e., alpha, beta, gamma, delta) primarily based totally to begin with on serology and later through phylogenetic clustering. To date, seven human coronaviruses had been constrained to the alpha and beta subgroups. An outbreak of an unknown breathing contamination in Wuhan China turned into mentioned in overdue December of 2019 and the causative agent turned into recognized as SARS coronavirus (SARS-CoV-2) and the sickness turned into referred to as coronavirus sickness 2019 (COVID-19). The sickness has hastily grow to be an international pandemic and a primary precedence has been located on locating capsules restrict viral propagation and infection (KIM et al, 2019).

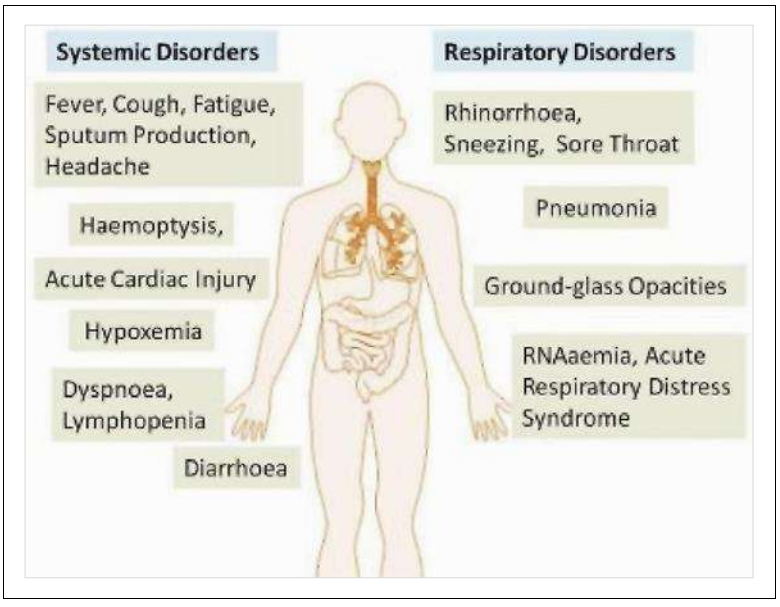


Figure 1. Systemic and Respiratory disorders of COVID-19 Rothan et al., 2020.

Turning to transportation methods, Person-to-person transmission happens primarily via direct contact or through droplets unfold by coughing or physiological reaction from Associate in Nursing infected individuals. Symptoms that appear, the signs of COVID-19 contamination seem after an incubation duration of about 5.2 days. The duration from the onset of COVID-19 signs to dying ranged from 6 to forty-one days with a mean of 14 days. This duration is depending on the age of the affected person and the repute of the affected person's immune system. It becomes shorter amongst patients >70-years antique in comparison with the ones below the age of 70 (HUSSIN et al, 2020).

According to the ICTV, human coronavirus belongs to the Betacoronavirus genus, a member of the Coronaviridae family, categorized in the order Nidovirales. It has been categorized into several genera, based on phylogenetic analyses and antigenic criteria, namely: (i) Alphacoronavirus, responsible for gastrointestinal disorders in humans, dogs, pigs and cats; (ii) Betacoronavirus, including the bat coronavirus, the human severe acute respiratory syndrome (SARS) virus, the Middle Eastern respiratory syndrome (MERS) virus and now the SARS-CoV-2; (iii) Gamma coronavirus, which infects avian species; and (iv) Deltacoronavirus.

Coronaviruses contain a positive-sense, single-stranded RNA genome. The genome size for coronaviruses ranges from 26.4 to 31.7 kilobases, and it is one of the largest among RNA virus (MACHADO et al, 2020).

Structurally, SARS-CoV-2 has four main structural proteins including spike (S) glycoprotein, small envelope (E) glycoprotein, membrane (M) glycoprotein, and nucleocapsid (N) protein, and also several accessory proteins (ASTUTI et al, 2020).

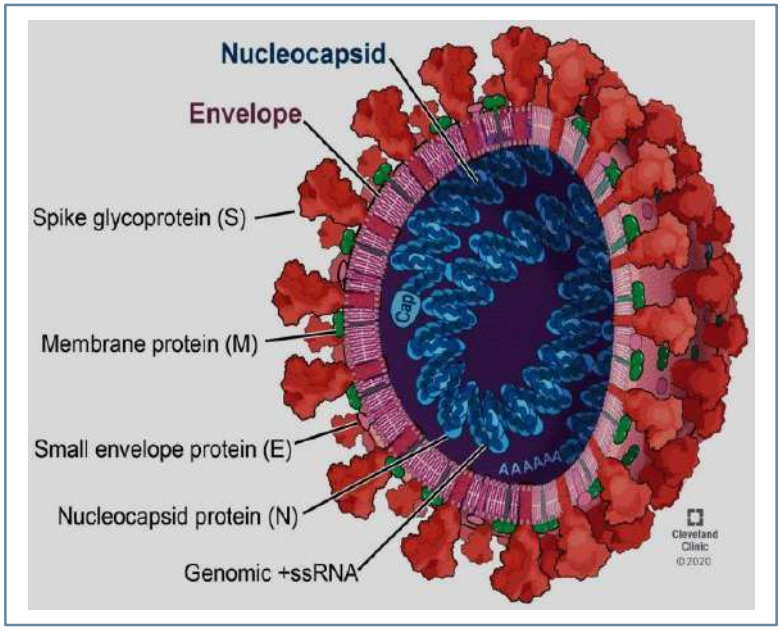


Figure 2. SARS-CoV-2 structural protein (BERGMAN et al, 2020).

The nucleocapsid known as N protein is the structural component of Corona virus localizing in the endoplasmic Reticulum-Golgi region that is structurally bound to the nucleic acid material of the virus. Because the protein is bound to RNA, the protein is involved in processes related to the viral genome, the viral replication cycle, and the cellular response of host cells to viral infections (ASTUTI et al, 2020).

The complete amino acid sequences of the nucleocapsid protein from the selected coronavirus representatives were compared to assess the level of similarity found between the conserved protein sequences of the classification-related viruses (BRUNO et al, 2020). He depicts the phylogenetic classification of corona viruses based on their NC protein. From this point of view, the RaTG13 bat virus ranks as the most similar to the circulating SARS-CoV-2, followed by SARS-Cov (BRUNO et al, 2020).

The below diagram (Figure 3) summarizes the main historical flow of coronavirus in what concerns the twenty-first-century epidemics (MACHADO et al, 2020).

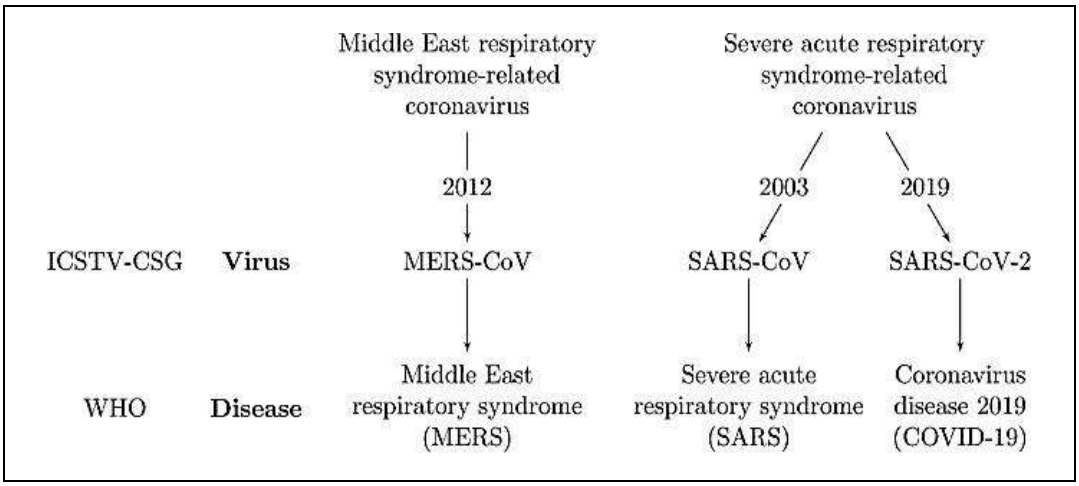


Figure 3. Historical flow of coronavirus

Given the continuing increase worldwide in cases of COVID-19, Four immune epitopes, S14P5, S20P2, S21P2 G and N4P5, were identified on the viral proteins S and N. IgG responses to all identified epitopes showed a strong detection profile, with N4P5 achieving the highest accuracy (100%) and sensitivity (> 96%) against SARS-CoV-2. Moreover, the magnitude of IgG responses to S14P5, S21P2, and N4P5 strongly correlated with disease severity (AMRUN et al, 2020).

Bioinformatics can appear to be a challenging field. Since it's a combination of the complex science of Biology with the complex theory of computer (HOLLOWAY et al, 2020). Bioinformatics's tool helps to design, compare, and forecast the structural and functional characteristics of proteins and genes. Also, Biology, Molecular biology, in particular, is undergoing two related transformations. First, there is a growing awareness of the computational nature of many biological processes and that computational and statistical model can be used to great benefit.

Second, developments in high-throughput data acquisition produce requirements for computational and statistical sophistication at each stage of the biological research pipeline (GENTLEMAN et al, 2004).

Moreover, Computational biology and bioinformatics have the potential to speed up the drug discovery process, thus reducing the costs, and to change the way drugs are designed (ALBERTO et al, 2006).

In our study, we used computational analysis to confirm the similarity of the SARS-COV-2 isolated around the world and exploring the evolutionary relationships with five other corona viruses of related taxonomical levels (BAT COV, SARS COV, MERS COV, Camel COV and Avian COV). Besides we compared the Nucleocapsid domains of *SARS-COV-2* with its related viruses and the configurations of the Receptor binding domain (RBD) with their specific antibodies.

In this context, The aims of this study were to prove the similarities between Corona viruses isolates (SARS COV-2) taken from different countries all over the world, compare between the whole genome sequences of SARS-COV-2 with taxonomically related corona viruses, computational comparative analysis of the nucleocapsid protein of SARS-COV-2 with taxonomically related corona viruses, alignment of sequences of SARS-COV-2 nucleocapsid protein with the sequences of the taxonomically related viruses, compare the structure of the receptor binding domain (RBD) of SARS-COV-2 with the 5 taxonomically related viruses and finally compare between the configurations of RBD of SARS-COV-2 with its specific antibody and with their taxonomically related viruses.

Materials and Methods

Materials

NCBI

The National Center for Biotechnology Information (NCBI), is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH). The NCBI is located in Bethesda, Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper (SMITH et al, 2013).

BLAST

BLAST Tool is an algorithm used for calculating sequence similarity between biological sequences such as nucleotide sequences of DNA and amino acid sequences of proteins. The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences (WHEELER et al, 2007). The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST

can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

Phyre2

Phyre2 is a pack of tools to predict and analyze protein structure, function and mutations. It has advanced remote homology detection methods to build 3D models, predict ligand binding sites and analyze the effect of amino acid variants (KELLEY et al, 2015). In the domain analysis fil, the prediction in the tool is 3 states: either α-helix, β-strand or coil. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coil.

The focus of Phyre2 is to provide biologists with a simple and intuitive interface to state-of-the-art protein bioinformatics tools. In this updated protocol, we describe Phyre2, which uses advanced remote homology detection methods to build 3D models, predict ligand binding sites and analyze the effect of amino acid variants. This protocol will guide users from submitting a protein sequence to interpret the secondary and tertiary structure of their models, their domain composition and model quality.

SCOPE

SCOPE is a manually curated ordering of domains from the majority of proteins of known structure in a hierarchy according to structural and evolutionary relationships (FOX et al, 2014).

PDB

The Protein Data Bank (PDB) was established as the 1st open access digital data resource in all of Biology and Medicine. Through an internet information portal and downloadable data archive, the PDB provides access to 3D structure data for large biological molecules (proteins, DNA, and RNA). Knowing the 3D structure of a biological macromolecule is essential for understanding its role in human and animal health and disease, its function in plants and food and energy production. The Vision of the RCSB PDB is to enable open access to the accumulating knowledge of 3D structure, function, and evolution of biological macromolecules, expanding the frontiers of fundamental biology, biomedicine, and biotechnology (HELEN et al, 2000).

Data Collection

Table 1. Showing the accession number of the viruses and complementary protein GI

Virus	Accession numbers	Protein GI
Bat CoV	MN996532	QHR63308.1
SARS-CoV	NC_004718.3	AAX16200.1
Camel CoV	MK967708	QGV13487.1
MERS-CoV	NC_019843.3	YP_009047211.1
Avian CoV	NC_001451	NP_040838.1

Method

Computational analysis

We used NCBI for assembling the samples at the beginning to compare the isolation of SARS-COV-2 from several countries all over the world to indicate the similarities between the different isolates.

The whole-genome sequences of SARS-CoV-2 were compared and analyzed with selected viruses of the Corona family: BAT-COV, SARS-COV, MERS-COV, CAMEL-COV, AVIAN-COV. for assessing the level of similarity between them by using Blast and that was the second step.

Then we compared between multiple sequence alignment of the nucleocapsid protein of SARS-COV-2 with related Coronavirus by using Nucleotide Blast. The nucleocapsid proteins sequences downloaded from the NCBI. Therefore, we have mentioned the accession number for each one as in Table 1.

After that, we collected the NC protein domain for SARS-COV-2 and the related Coronavirus domains as in the table 2. Where we compared the nucleocapsid structures of the domains. and the tool that we used for all of this was phyre2.

By using SCOPE, we Compared between configurations, the number of amino acids, and the diameters of the domains of the NC proteins of SARS-COV-2 with the related Coronaviruses.

Finally, we determined the configurations of the receptor binding domain RBD with 3D-specific antibody for SARS COV2 and compared it with: Bat-COV, SARS-COV, Camel-COV, MERS-COV and Avian-COV.

Results

1. Alignment between SARS-COV-2 and other strains isolated all over the world

We proved in these results a great similarity between SARS-COV-2 emerging from Wuhan, China and other strains isolated all over the world. The percentage similarity is ranging from (99.98%-100.00%), This means that the new SARS-COV-2 isolate emerging from Wuhan, China is circulating around the world (Table 2).

Table 2. SARS-COV-2 isolates from 12 different countries around the world

Isolates definition	Isolated from	Identity	Accession ID
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/POL/PL_P11/2020, complete genome	Poland	99.98%	MT511067.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/howo sapiens/MAR/RMPS05/2020, complete genome	Morocco	99.98%	MT731746.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ESP/HUD-70062002/2020, complete genome	Spain	99.98%	MT956918.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/URY/Mdeo-1/2020, complete genome	Uruguay	99.99%	MT466071.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/DEU/FFM5/2020, complete genome	Germany	99.99%	MT358641.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IND/GMCKN318/2020, complete genome	India	99.99%	MT415321.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/VNM/nCoV-19-01S/2020, complete genome	Vietnam	100.00%	MT192772.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/BRA/RJ-DCVN2/2020, complete genome	Brazil	99.99%	MT835026.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TWN/CGMH- CGU-09/2020, complete genome	Taiwan	99.99%	MT374105.1
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/France/40002VJ/2020, complete genome	France	99.98%	MT470179.1
Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome	Wuhan	100.00%	NC_045512.2
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB1763/2020, complete genome	USA	100.00%	MT671817.1

2. Alignment of the whole genome of SARS-COV-2 with the five taxonomically related corona viruses

By comparing the sequences of the whole genome of SARS-COV-2 with the 5 taxonomically related Corona viruses using BLAST tool, we found that BAT-COV is the most similar to SARS-CoV-2 96%, followed by SARS-COV 82%, then MERS- COV and CAMEL-COV 67%, the Avian-COV is featured by the lowest similarity with SARS-CoV-2 65% (Table 3).

Table 3. The whole genome sequences of SARS-COV-2 compared with the other selected viruses to assess the level of similarity between them

Name of the Virus	Accession No	Name of the comparable virus	Identities	Accession No
SARS- COV-2 complete genome.	NC_045512	Recombinant coronavirus clone **Bat-CoV**, complete sequence	96%	MN996532
		SARS coronavirus CUHK-W1, complete genome	82%	NC_004718.3
		Middle East respiratory syndrome coronavirus isolate **MERS-CoV**-Jeddah-human-1, complete genome	67%	NC_019843.3
		Camel alphacoronavirus Camel229E isolate Camel229E-CoV/JC50_HUH7/KSA/2014, complete genome	67%	MK967708
		Avian coronavirus isolate CoV-9698-2016/11/28-BP/SA orf1ab polyprotein (orf1ab) gene, partial cds	65%	NC_001451

3. Alignment of sequences of SARS-COV-2 nucleocapsid protein with the sequences of the taxonomically related viruses

By comparing the amino acid sequences of the nucleocapsid proteins (NC) of SARS-COV-2 with the sequences of the five taxonomically related viruses, we found that the NC of Bat-COV was almost identical to SARS- COV-2 with 98%. Then followed by SARS-COV, with 90%. While MERS-COV and Camel COV, display 48% and that was far away from Bat COV ratio similarity. At last, Avian COV clearly does not match where it displays only 29% (Table 4).

Table 4. Comparison the NCP of SARS-COV-2 with its related coronaviruses

Name of the virus	Protein GI of the nucleocapsid	Identity
Bat SARS-COV	QHR63308.1	98%
SARS – COV	AAX16200.1	90%
CAMEL COV	QGV13487.1	48%
MERS – COV	YP_009047211.1	48%
AVIAN COV	NP_040838.1	29%

It is clear that this result matches with the previous result that BAT-COV is the most similar virus to SARS-COV-2 followed by SARS-COV.

4. Comparative computational analysis of the nucleocapsid (NC) protein of SARS-COV-2 with taxonomically related corona viruses

The receptor binding domains (RBD) of the nucleocapsid proteins of SARS-COV-2 was compared with those of the 5 taxonomically related viruses using Phyre 2. The structure of the RBD both SARS-COV and Bat-COV (d2cjra1) are identical to the novel virus (SARS-COV-2) and conserved in it (Fig. 4), The structure of the RBD of both MERS-COV and Camel-COV (d2giba1) are identical and conserved (Fig. 5) while the Structure of the RBD of Avian-COV is different Fig. 6 and Table 5.

Table 5. Structural domains of SARS-COV-2 NC protein and its related coronaviruses

Virus	Domain	Alpha helix	ratio of alpha	Beta strand	ratio of Beta	Coils
SARS COV 2	d2cjra1	11	22%	6	10%	24
BAT COV	d2cjra1	11	22%	6	10%	24
SARS-COV	d2cjra1	11	22%	6	9%	24
CAMEL COV	d2giba1	9	23%	6	14%	18
MERS COV	d2giba1	9	23%	6	14%	18
AVIAN COV	d2ca1a1	11	19%	6	14%	20

Shapes of the receptor binding domains:

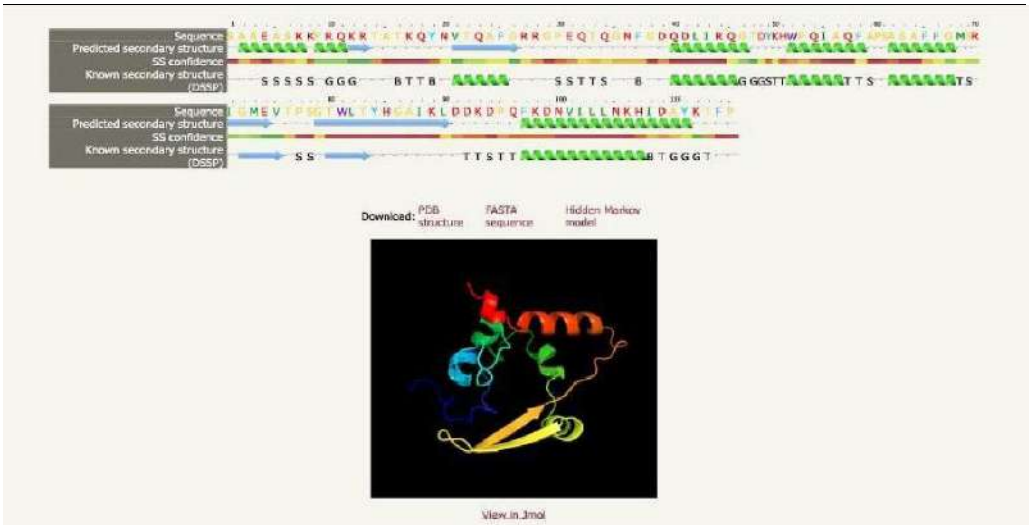


Figure 4. RBD (d2cjra1) conserved in SARS-COV- 2, BAT-COV, SARS-COV

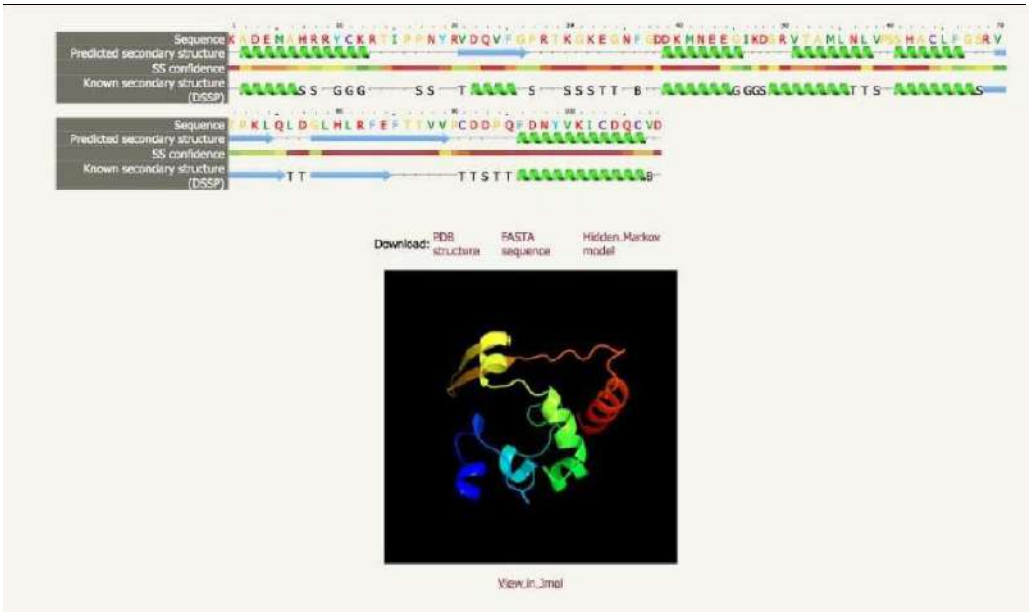


Figure 5. RBD (d2giba1) conserved in MERS-COV and CAMEL- COV

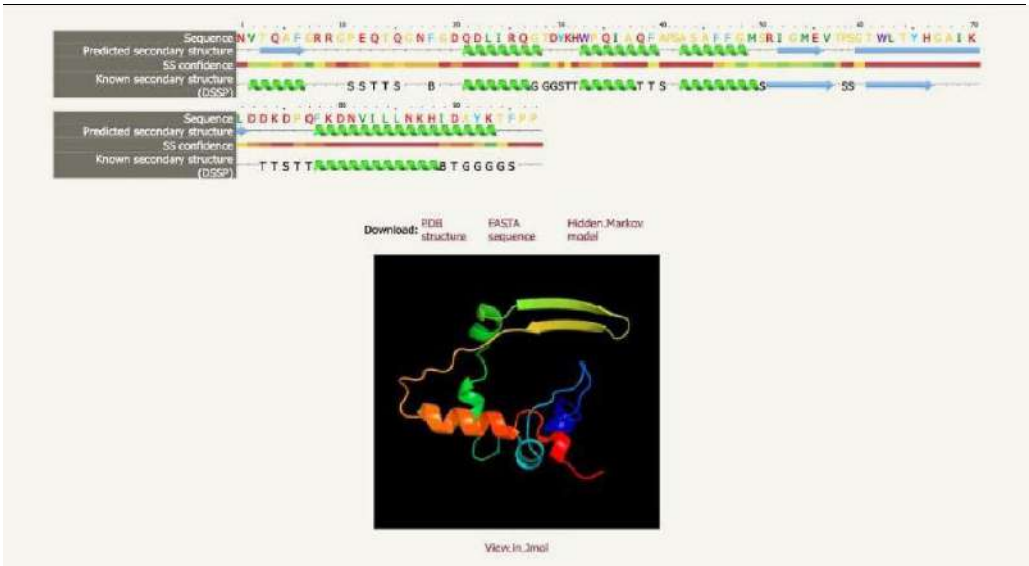


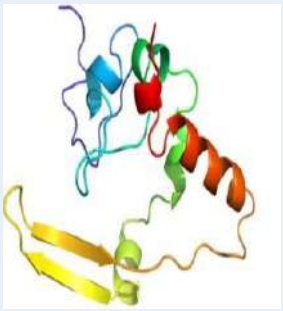
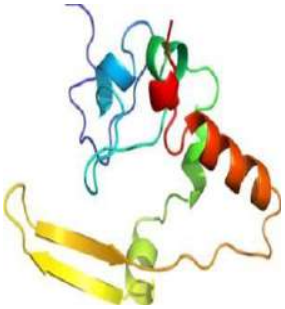
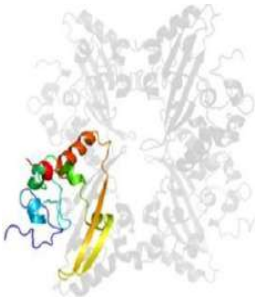
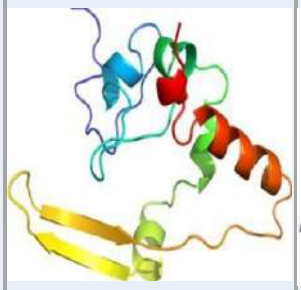
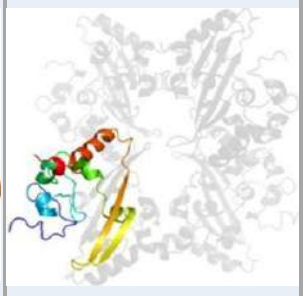
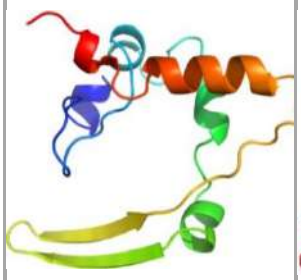
Figure 6. RBD (d2ca1a1) found in Avian-COV

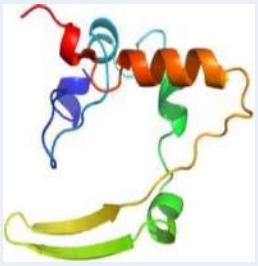
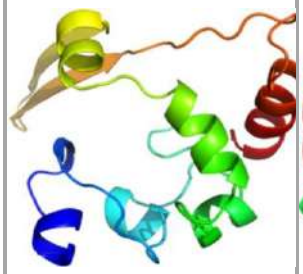
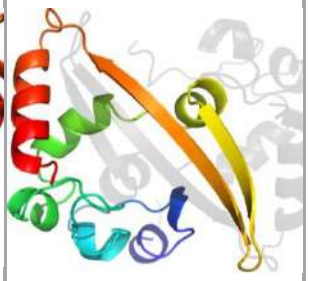
5. Compare the structure of the receptor binding domain (RBD) of SARS-COV-2 with the 5 taxonomically related corona viruses.

We compared the structure of the RBD of the NC proteins of SARS-COV-2 with the other 5 taxonomically related viruses by using SCOPE tool that determine the length of the domain, diameter and its configuration as illustrated in Table 6.

We found that the RBD of SARS-COV-2 is nearly similar to BAT–COV and SARS-COV while MERS-COV, Camel-COV and Avian-COV are different.

Table 6. Comparison between configurations, no. of amino acids and diameter of the domains

Name	Domain	FASTA sequence	length	Diameter	Shape 0f the domain	Configuration of the nucleocapsid
SARS COV2	**d2cjra1**	SAAEASKKPRQK RTATKQYNVTQA FGRRGPEQTQGN FGDQDLIRQGTD YKHWPQIAQFAP SASAFFGMSRIGM EVTPSGTWLTYH GAIKLDDKDPQF KDNVILLNKHIDA YKTFP	**365-251 114**	**2.5A**		
Bat-COV	**d2cjra1**	SAAEASKKPRQK RTATKQYNVTQA FGRRGPEQTQGN FGDQDLIRQGTD YKHWPQIAQFAP SASAFFGMSRIGM EVTPSGTWLTYH GAIKLDDKDPQF KDNVILLNKHIDA YKTFP	**365-251 114**	**2.5A**		
SARS- COV	**d2cjra1**	SAAEASKKPRQK RTATKQYNVTQA FGRRGPEQTQGN FGDQDLIRQGTD YKHWPQIAQFAP SASAFFGMSRIGM EVTPSGTWLTYH GAIKLDDKDPQF KDNVILLNKHIDA YKTFP	**365-251 114**	**2.5A**		
MERS COV	**d2giba1**	NVTQAFGRRGPE QTQGNFGDQDLI RQGTDYKHWPQI AQFAPSASAFFG MSRIGMEVTPSGT WLTYHGAIKLDD KDPQFKDNVILLN KHIDAYKTFPP	**366-270 96**	**1.75A**		

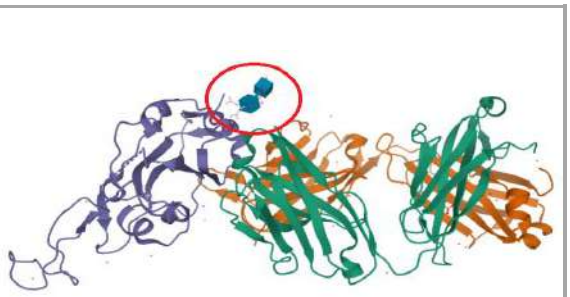
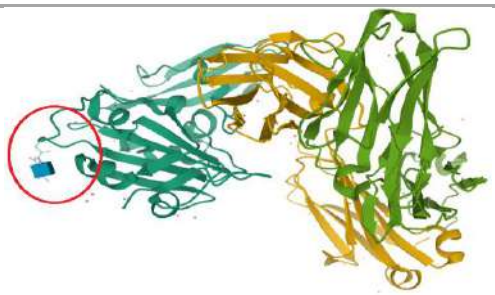
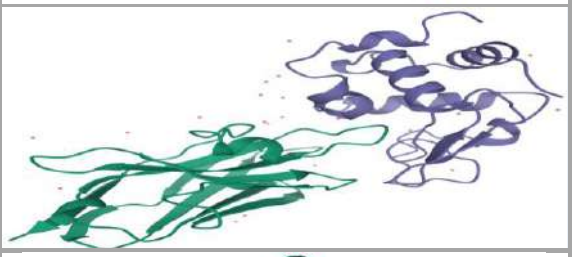
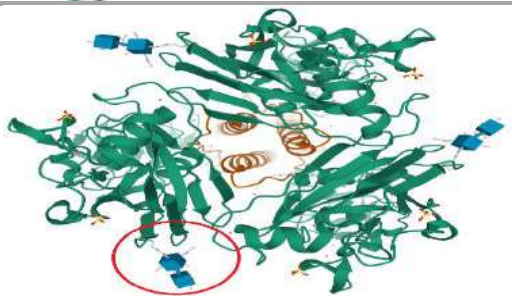
Camel COV	**d2giba1**	NVTQAFGRRGPE QTQGNFGDQDLI RQGTDYKHWPQI AQFAPSASAFFG MSRIGMEVTPSGT WLTYHGAIKLDD KDPQFKDNVILLN KHIDAYKTFPP	**366-270 96**	**1.75A**		
Avian COV	**d2ca1a1**	KADEMAHRRYC KRTIPPNYRVDQV FGPRTKGKEGNF GDDKMNEEGIKD GRVTAMLNLVPS SHACLFGSRVTPK LQLDGLHLRFEFT TVVPCDDPQFDN YVKICDQCVD	**326-218 108**	**2.6A**		

6. Compare between the configurations of RBD of SARS-COV-2 with its specific antibody and with their related corona viruses.

We used RCSB PDB for comparing the configurations of binding the RBD with its specific antibodies (Table 7). We found that although the RBD is similar to some viruses, the way it binds to antibodies is different, and this explains the difference in the immune response from one virus to another.

Table 7. Comparison between SARS-COV-2 and related coronaviruses configurations of the RBD with the specific antibody

Name of the Virus	RBD (PDB ID)	Configuration of RBD with specific antibody
SARS-COV-2	7JMO	
BAT-COV	5GYQ	

SARS-COV	7JN5	
MERS-COV	5DO2	
CAMEL-COV	1MEL	
AVIAN-COV	3S13	

Discussion

The current health emergency raised by the COVID-19 aroused the interest of the world-wide scientific community that actively operates under diverse research lines on the attempt to address the multiple facets of this pandemic issue, including the elucidation of the immunological peculiarities of the novel virus.

In our study, the reference sequence for SARS-CoV-2 was constructed based on genomic sequence of 63 isolated strains. The genome size of the reference strain was 29870 bp. The reference sequence was identical to the genomic sequence of 15 strains (15.79%) isolated from clinical samples, suggesting the reference strain would display full biological functions and pathogenicity (WANG et al, 2020).

As a typical RNA virus, the evolution rate of coronavirus could be substituted per base pair per year, and mutation could occur during each replication cycle. However, phylogenetic analysis and sequence alignment showed that the homology among different isolates was extremely high. Compared to the reference sequence, the homology of the vast majority of isolates was above 99.99% at both the nucleotide and amino acid levels. In fact, the homology in most encoding regions was 100%. Overall, results from our analyses suggested that the virus in this epidemic might originate from the same animal species, and caused widespread infection in a short period of time (HUANG et al, 2020).

In our study, the genome sequences of the selected viruses were verified (Table 3). We found that BAT-CoV is one of the relatives of SARS-CoV groups. This corona virus is the closest known relative to SARS-CoV-2 with 96% identity followed by SARS-COV (82%), and MERS-COV (67%). These findings match with (FANI et al, 2020) who confirmed that SARS-CoV-2 is 96.2% identical to some bat coronaviruses and also more distantly correlated to SARS-CoV and MERS-CoV (about 79 and 50%, respectively).

Nucleocapsid protein is a highly conserved protein involved in the active phase of the viral infection. Being highly immunogenic is commonly targeted in studies aimed at developing alternative diagnostic tools and prophylactic strategies (CHANG et al, 2004, AHMED et al, 2020, ZHANG et al, 2020 GRIFONI et al, 2020 and BRUNO et al, 2020).

By comparing the nucleocapsid (NC) protein of SARS-COV-2 with the five taxonomically related coronaviruses (table 4), we found that NC of BAT-COV is closely related to SARS-COV-2 99% followed by SARS- COV and MERS COV 90% and 48% respectively Whereas, the Avian corona virus was featured by the lowest similarity with the circulating SARS-CoV-2 and this match with BRUNO et al., 2020.

With regard to epitopes and domains distribution across viruses with tropism for domestic animals, the survey highlights that some domains sequences are more "conserved" among the most related specimens, while other domains are shared among a wider ensemble of coronaviruses. Similar outcomes were also described for the spike proteins epitopes BRUNO et al., 2020. In this view, a detailed description of the domains and epitopes distribution over the viral population might provide valuable information driving future researches aimed at setting efficient prophylactic strategies and/or the design of tool capable of differential diagnosis on the basis of serological tests.

In this study, we found that the RBD of SARS-COV-2 is nearly similar to BAT–COV and SARS-COV while MERS-COV, Camel-COV and Avian-COV are different (Tables 5, 6), additionally, we found that the RBD of both MERS-COV and camel-COV are similar, this means that the sequence of RBD are more "conserved" among the most related specimens.

By using RCSB PDB, we found that although the RBD is closely similar to some closely related viruses, the way it binds to antibodies is different (Table 7), and thus explains the difference in the host immune response from one virus to another. Thus, potentially involved in conferring partial protection against SARS-CoV-2 infection other being considered good candidates for the development of diagnostic tools and prophylactic-oriented strategies.

Conclusions

SARS-COV-2 in this epidemic might originate from bats and caused widespread infection. BAT-COV is the most similar to SARS-CoV-2 (96%), followed by SARS-COV (82%). The nucleocapsid of BAT-COV is closely related to SARS-COV-2 followed by SARS-COV and MERS COV respectively. Whereas, the Avian-COV was featured by the lowest similarity. Although the receptor binding domain is similar in some closely related viruses, the way it binds to antibodies is different thus explains the difference in the host immune response from one virus to another.

To conclude, COVID-19 pandemic is a global issue that stimulates the synergistic cooperation between research groups all over the world. At the same manner, planning studies that seek to address multiple aspects of COVID-19 enable a fast and effective control of the SARS-CoV-2 infections. In line with the One-Health concept, extending the target of the studies to the environment and the animals which humans are continuously interacting with favor a better understanding of this complex phenomenon and help to adopt the most suitable containment measures.

Continuing mapping the major epitopes and domains known for the viral proteins is, in our view, a promising strategy. In this light, defining the map of the circulating epitopes and domains among diverse geographical areas and considering the diverse people habits and animal population they are interacting with might also provide valuable information for the design of personalized prophylactic and diagnostic strategies, shaped in dependence of the geographical context and the subject's habits. Thus, an incisive focus on these aspects can contribute to a relatively fast control of this pandemic issue.

Conflict of Interest

The author has no conflict of interest to declare.

Acknowledgments

This project was supported by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University.

References

1. ADHIKARI S., S. MENG, Y. WU. (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review 3:17-25.
2. AHMED S.F., QUADEER A.A., MCKAY M.R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV *immunological studies.* Viruses 12:254-260.
3. ALBERTO AMBESI-IMPIOMBATO and DIEGO D. BERNARDO (2006). Computational Biology and Drug Discovery: From Single-Target to Network Drugs", *Current Bioinformatics* 1: 3-10.
4. AMRUN SITI NAQIAH, CHERYL YI-PIN LEE ,BERNETT (2020). Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *RESEARCH PAPER* 58: 102911-102920.
5. ASTUTI, INDWIANI, and YSRAFIL (2020). "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response." *Diabetes & metabolic syndrome* 14(4): 407-412.
6. BERGMANN CORNELIA C., and SILVERMAN ROBERT H. (2020). COVID-19: Coronavirus replication, pathogenesis, and therapeutic strategies. *Cleveland Clinic Journal of Medicine* 87 (6): 321-327.
7. BRUNO TILOCCA, ALESSIO SOGGIU B., MAURIZIO SANGUINETTI (2020). Comparative Computational Analysis of SARS-CoV-2 Nucleocapsid Protein Epitopes in Taxonomically Related Coronaviruses." *Microbes and Infection* 22.4-5: 188-194.
8. CHANG M.S., LU Y.T., HO S.T., WU C.C., WEI T.Y., CHEN C.J. (2004). Antibody detection of SARS-CoV spike and nucleocapsid protein. *Biochem Biophys Res Commun* 314:931-936.
9. CHINSEMBU, K.C. (2019). Coronaviruses and Nature's Pharmacy for the Relief of Coronavirus Disease. *Rev. Bras. Farmacogn*. 30:603-621.
10. FANI, MONA (2020). Comparison of the COVID-2019 (SARS-CoV-2) pathogenesis with SARS-CoV and MERS-CoV infections. *Future Virology* 15(5): 1746-0794.
11. FOX NK, BRENNER S.E., CHANDONIA J.M. (2014) "SCOPE: Structural Classification of Proteins extended, integrating SCOP and ASTRAL data and classification of new structures." *Nucleic acids research* 42: 304-309.
12. GENTLEMAN, R.C., CAREY, V.J., BATES, D.M. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol 5*, R80, Article number: R80.
13. GRIFONI A., SIDNEY J., ZHANG Y., SCHEUERMANN R.H., PETERS B., SETTE A. (2020). A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell host microbe.* [Online ahead of print].
14. HELEN M. BERMAN, JOHN WESTBROOK, ZUKANG FENG, GARY GILLILAND, T.N. BHAT, HELGE WEISSIG, ILYA N. SHINDYALOV, PHILIP E. BOURNE (2000). *The Protein Data Bank, Nucleic Acids Research*, 28(1): 235-242.
15. HOLLOWAY ERIC (2020) Tutorial: Bioinformatics Basics, Communications of the Blyth Institute 35-38.
16. HUANG C, WANG Y, LI X. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *ARTICLES*| 395 (10223): 497-506.
17. HUSSIN A. ROTHAN, SIDDAPPA N. BYRA-REDDY (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, *Journal of Autoimmunity* 109: 0896-8411.
18. KELLEY, L., MEZULIS, S., YATES, C. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10, 845-858.
19. KIM, J., ZHANG, J., CHA, Y. (2020). Advanced bioinformatics rapidly identifies existing therapeutics for patients with coronavirus disease-2019 (COVID-19). *J Transl Med* 18, 257.
20. MACHADO, J.A.T., ROCHA-NEVES, J.M. & ANDRADE, J.P. (2020). Computational analysis of the SARS-CoV-2 and other viruses based on the Kolmogorov's complexity and Shannon's information theories. *Nonlinear Dyn*: 1731-1750.
21. SMITH K. (2013). A Brief History of NCBI's Formation and Growth. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): *National Center for Biotechnology Information* (US).
22. XIE, M & CHEN, Q. (2020). Insight into 2019 novel coronavirus – An updated interim review and lessons from SARS-CoV and MERS- CoV. *International journal of infectious diseases; 94*: 119-124.
23. WANG C, LIU Z, CHEN Z et al. (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. *Journal of medical virology* 92(6):667-674.
24. WHEELER D, BHAGWAT M. (2007). BLAST QuickStart: Example-Driven Web-Based BLAST Tutorial. *Methods in molecular biology* 395:149-176.
25. ZHANG P., GAO Q., WANG T., KE Y., MO F., JIA R. (2019). Evaluation of recombinant nucleocapsid and spike proteins for serological diagnosis of novel coronavirus disease (COVID-19) *MedRxiv*. doi: 10.1101/2020.03.17.20036954
26. ZHONGHUA LIU XING BING XUE ZA ZHI. (2020) Feb. Epidemiology Working Group for NCIP Epidemic Response, *Chinese Center for Disease Control and Prevention*. 10; 41(2):145-151.